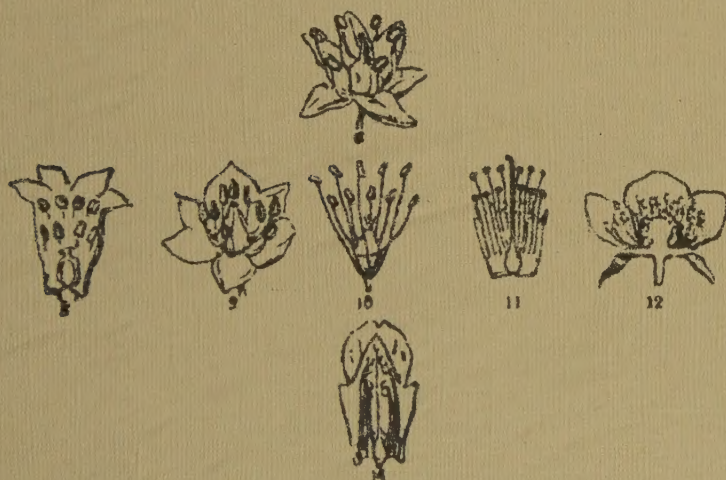


ON THE USE OF PRINCIPAL COMPONENTS AND FACTOR ANALYSIS IN ECOLOGY

SOME APPLICATIONS OF ORDINATION
TECHNIQUES TO VEGETATIONAL,
ENVIRONMENTAL AND MICROBIAL DATA



EWA KVILLNER
DEPARTMENT OF PLANT ECOLOGY
UNIVERSITY OF LUND, SWEDEN

Lund 1978

ON THE USE OF PRINCIPAL COMPONENTS AND FACTOR ANALYSIS
IN ECOLOGY

Some applications of ordination techniques to vegetational,
environmental and microbial data

Ewa Kvillner

Fil mag, Kr.

Akademisk avhandling som för avläggande av filosofie
doktorsexamen vid matematisk-naturvetenskapliga fakul-
teten vid universitetet i Lund kommer att offentligen
försvaras å Institutionen för ekologisk botanik, Östra
Vallgatan 14, Lund, fredagen den 17 februari 1978
kl 10.00.



"Att söka med fullständig noggrannhet känna de mångfaldiga former, hvilka växtriket öfver hela jordens yta företer, att utfinna och förklara de medel, genom hvilka, och de organer, hvarest lifvet verkar i hvarje planta, och att sedermera till människans sanna nytta lämpa de vunna kunskaperna om växtens bildningar och egenskaper - detta är botanikens uppgift.

Hvarje ordnande af naturalstren, hvarje förening af de särskilda släktena till familjer, ordningar och klasser, kan ske på tvenne sätt. Antingen bestämmes detta sammanhang efter blott ett enda eller få kännetecken; eller ock grundar man detsamma på öfverensstämmelse i de flesta eller väsentligaste delarna. Det förra kallas ett *konstmässigt*, *artificielt* system, det sednare ett *naturligt*. Så är det t ex ett konstmässigt system, om gossarna i en skola skulle indelas i klasser efter klädernas färg, om djuren skulle ordnas efter blott rörelseorganens beskaffenhet, eller växterna fördelas blott efter antalet af de ståndare, som finnas i hvarje blomma. På detta sätt blefve det visserligen temligen lätt att snart och säkert bestämma till hvad klass det ifrågavarande naturföremålet hör; men någon hel insigt i dess natur vinner ej genom denna bestämning. *Naturligt* blir systemet eller ordnandet deremot, dvs angifvande föremålets verkliga natur, om gossarne t ex fördelas i klasser efter deras själsutbildning, ådagalagd genom deras kunskaper och om djuren och växterna ordnas efter utvecklingen och öfverensstämmelsen af alla eller de flesta yttre och inre organerna.

För att förvärfva sig en riktig och gagnande insigt uti växternas former och om växtverldens skiftande mångfald äro botaniska utflygter eller exkursioner uti den fria naturen, på ångar och gården, i skog och mark, vid sjö och å, alldeles oundgängliga. Ty det är endast genom noggranna och flitiga iakttagelser på de lefvande växterna, genom ett förtroget umgänge med hela den yttre skapelsen, som kärleken till den kan vinnas och insigter om dess företeelser beredas."

Ur Inledning till botaniken af N.J. Andersson, 1871.

This thesis is based on the following four papers:

- I. Kvillner, E. and Sonesson, M. 1977. Ordination of plant distribution and environmental data. - Submitted to Sonesson, M. (ed.). Ecology of a Subarctic Mire. Ecological Bulletins, Vol. 27 (1978).
- II. Andersson, F. and Kvillner, E. 1977. Ecological studies in a Scanien woodland and meadow area, Southern Sweden. IV. Gradient analysis of vegetational and environmental patterns.
- III. Rosswall, T. and Kvillner, E. 1977. Principal components and factor analysis for the description of microbial populations. - Submitted to Advances in Microbial Ecology Vol. 2 (1978).
- IV. Kvillner, E. 1977. Applicability and validity of some numerical methods in plant ecology. I. Principal components analysis.

INTRODUCTION

During the last decades the application of numerical methods in ecological contexts has generated two different definitions for ordination - a summarizing term for a host of procedures among which principal components analysis and factor analysis are the most frequently used. Goodall (1954) proposed "ordination" as an antonym to "classification", indicating an ordering of relationships among items on a continuous scale as opposed to assigning items to discrete classes. Therefore a broad definition of ordination as "any arrangement summarizing relationships among items" has been suggested by, e.g., Austin (1972), Nichols (1977). A more restricted definition as "an arrangement of samples or species on axes relating to environmental gradients" has been developed with considerations to the traditional aims in plant sociology (Groenewoud, 1965, Swan, 1970). The current discussion on ordination techniques arises from the assumptions regarding "the ecological space" (Austin, 1976), which specifies the distribution of a population response to an ecological gradient. This is essential as the aim of principal components analysis and factor analysis is to reveal those relationships

between organisms and environment which are most likely to repay closer investigations.

The performances of ordination methods cannot be adequately assessed without considering the intended purpose of the methods and what constitutes a model of ecological variables against which these methods may be compared.

The aim of this work is to consider:

1. The inherent characteristics of principal components analysis, PCA, and factor analysis, FA.
2. Comparisons of vegetational, environmental and microbial data when applied to ordination techniques.
3. Relationship between the vegetation and the environment systems evaluated from PCA-ordinations on the separate systems and from canonical correlation analysis on both systems.
4. The use of correlation and covariance coefficients on species data.

Applications of numerical methods to ecological data considered in this dissertation

Material	Sampling method	Variables studied	Numerical methods
A nutrient-deficient mire complex on the elevated areas of Stordalen, Northern Sweden. The "720"-material (720 frames, each with 30 pins) and the "72"-material (72 quadrats, 2.5x2.5 m, each with 10 frames) (paper I)	point-frequency method	36 species occurring in the field- and bottom layer 4 environmental variables (peat water content, slope of surface, temperature at 10 cm depth, depth to permafrost)	PCA-ordination of correlation matrix on vegetation data simple correlations of environmental variables versus principal axes of the vegetation system
A deciduous woodland and meadow area of Linnebjerg, Southern Sweden. The "182"- and the "97"-materials (182 and 97 quadrats, 10x10 m) (paper II)	contiguous quadrat method	62 species occurring in the field layer 14 environmental variables of the humus layer (soil moisture, acid-base status, exchangeable cations, organic matter, soil structure)	PCA-ordination of covariance matrix on vegetation data PCA-ordination of correlation matrix on environment data simple correlations of environmental variables versus principal axes of the vegetation system canonical correlations of vegetation and environment systems
Microbial populations in terms of a number of individual isolates (paper III)	multipoints techniques	Test results of measurements on processes, e.g., litter decomposition, nitrogen fixation	PCA- and FA-ordinations of correlation matrices on microbial data

DISCUSSION

Principal components analysis reduces a set of variables (e.g., species, environmental variables or microbial tests) to a smaller subset (principal variables or axes) containing the same or almost the same amount of information. Factor analysis has a somewhat similar end but a different mathematical model (paper III). Both the ordination techniques are mathematically based upon assumptions of linearity, i.e., the relationships among organisms and environment are assumed to be linear. From an ecological point of view, however, it is assumed that the distributions of species populations may be represented by a series of overlapping bellshaped curves along a single environmental gradient (Whittaker, 1967). In relation to two or more gradients, the distributions of species populations become bellshaped response surfaces in a hyperspace. For the case of many variables, visualization of the surfaces will be evidently difficult. Mathematically, however, these problems can be solved. The principal axes of the ordination procedures are constructed in the centre of the hyperspace in such a way that they should account for as much of the total variance as possible. In the new coordinate system, the original straight line of an ecological gradient may become complex curves so that the least associated species come to lie closely adjacent, the so-called "horse-shoe effect". This effect and the degree of distortion is considered in terms of the strong dependence upon different variables used (paper IV).

Non-numerically recognized vegetational and environmental gradients in the Stordalen and Linnebjær materials have been compared with those revealed by principal components analysis. One of the major vegetational gradients in the mire vegetation, viz., the "hummock--mud-bottom" gradient is fairly easily followed along the first principal axes of the Stordalen "72"-material. This difference may be attributable to the different sample sizes used, revealing the

"horse-shoe effect" and the great heterogeneity in the "720"-material. The distribution of species along the second principal axes resembled to some extent that along the large-scale vegetational "mire margin-mire expanse" gradient. The primary relationship between vegetation and environment was found in terms of variation of the water content and the microtopography (paper I).

The major vegetational gradient "low herb - tall herb", dividing the species into woodland and meadow, is found along the diagonal of the first two axes in the PCA-ordination of the Linnebjerg "97"- and "182"-materials (paper II). The ecological gradient "dry-wet", non-numerically constructed from consideration of the variation in soil moisture and soil aeration was distorted according to the "horse-shoe"-effect. An ecological gradient, numerically recognized in the environment data from considerations of acid-base status, soil moisture and stores of exchangeable cations, is comparatively less distorted. Much the same gradient was found by use of canonical correlation analysis on variables in both the vegetation and the environment systems. The vegetational gradient "tall herb-sedge", non-numerically recognized in the open meadow vegetation, is but partly found on the principal axes of the PCA-ordination. However, some of the distortions of that gradient may be due to the large sample size. Therefore, part of the important variation is hidden and the degree of resolution is decreased.

Thus, the non-numerical and numerical gradient analyses gave similar results. However, absence of a species affected the numerical analysis more than the non-numerical analysis. The observations in the field as they were expressed in the non-numerical descriptions formed a hypothesis of the relation between the vegetational and environmental features. The numerical analyses were then used to test and refine the hypothesis. Obvious similarities were found in the results and to a great extent the hypothesis was validated. The

numerical analysis also suggested further hypotheses.

For quantitative as well as for qualitative species data the relationship between species variance(s^2) and mean frequency or per cent presence seems to be curvilinear with a peak at 50% mean frequency/presence (paper IV). This should be kept in mind when decisions on proper coefficients measuring species interrelationships are made. The correlation coefficient of Pearson removes the dependence of the variance on the mean. Instead rare species will have a significant influence on the development of the principal axes of PCA-ordination. Species positioned near the origin in a PCA-ordination based on a covariance matrix may have an important ecological meaning but their variation is overwhelmed by other species. Such situations may not always be desirable and both frequent and infrequent species must be considered. For all data transformations special care must be taken not to overwhelm the PCA-analysis by purely numerical considerations.

The degree of presence of a species can always be recorded, but there is no corresponding measure of the degree of absence (see however Swan, 1970). A large number of zero values, i.e., absence of a species, will strongly affect the PCA-ordination (like any other numerical method) while environmental variables can be recorded along the whole of the ecological gradient. They are usually more or less normally distributed or at least they can be transformed to, e.g., a logarithmic scale if the distribution on linear scale is skew. Therefore, the use of ordination techniques on environmental data is more likely to be "satisfactory" (paper IV).

A further application of the ordination techniques was performed by PCA and FA on microbial data. The tests results give a description of the physiological qualities of the microbial populations. The latter are often hazardous to describe and the microbial study is more often oriented on

process than on structure. Therefore, the reducing power and the generation of hypotheses on functional relationships seem to be the most important advantages when considering the use of principal components and factor analysis on microbial data (paper III). Application of the two ordination techniques in environmental microbiology seems to be useful in evaluating the effect of a wide range of stress factors on populations of bacteria, e.g., acid precipitation, eutrophication and manure decomposition.

CONCLUDING REMARKS

Despite the effect of assumptions of linearity, principal components and factor analysis are still two exploratory techniques of great value especially in the initial stages of an ecological study. The reducing power and the generation of hypotheses regarding causal relationships among organisms and environment may be the most important advantages of the procedures. The broad definition of ordination as "any arrangement summarizing relationships among items" seems to be valid for the two techniques used.



REFERENCES

- Austin, M.P. 1972. Models and analysis of descriptive vegetation data. - In: J.N.R. Jeffers (ed.) *Mathematical Models in Ecology*. Symp. Brit. Ecol. Soc.:12:61-86.
- Austin, M.P. 1976. On non-linear species response models in ordination. - *Vegetatio*, Vol 33, 1:33-41.
- Goodall, D.W. 1954. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. - *Aust. J. Bot.* 2:304-324.
- Groenewoud, H. van. 1965. Ordination and classification of Swiss and Canadian coniferous forests by various biometric and other methods. - *Ber. Geobot. Forsch. Inst. Rübel* 36:28-102.
- Nichols, S. 1977. On the interpretation of principal components analysis in ecological contexts. - *Vegetatio*, Vol 34, 3:191-197.
- Swan, J.M.A. 1970. An examination of some ordination problems by use of simulated vegetation data. - *Ecology* 51:59-102.
- Whittaker, R.H. 1967. Gradient analysis of vegetation. - *Biol. Rev.* 49:208-264.

